

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Rasmussen, Grethe
Mikkelsen, Jan Moller
Schulein, Martin
Patkar, Shankant A.
Hagen, Fred
- (ii) TITLE OF INVENTION: A Cellulase Preparation Comprising an
Endoglucanase Enzyme
- (iii) NUMBER OF SEQUENCES: 33
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Novo Nordisk of North America, Inc.
 - (B) STREET: 405 Lexington Avenue, 64th Floor
 - (C) CITY: New York
 - (D) STATE: New York
 - (E) COUNTRY: United States of America
 - (F) ZIP: 10174-6401
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/389,423
 - (B) FILING DATE: 14-FEB-1995
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Lambiris, Elias J.
 - (B) REGISTRATION NUMBER: 33,728
 - (C) REFERENCE/DOCKET NUMBER: 3469.214-US
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 212-867-0123
 - (B) TELEFAX: 212-878-9655

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1060 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Humicola insolens*
 (B) STRAIN: DSM 1800

(ix) FEATURE:

- (A) NAME/KEY: mat_peptide
 (B) LOCATION: 73..924

(ix) FEATURE:

- (A) NAME/KEY: sig_peptide
 (B) LOCATION: 10..72

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION: 10..924

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGATCCAAG ATG CGT TCC TCC CCC CTC CTC CCG TCC GCC GTT GTG GCC	48
Met Arg Ser Ser Pro Leu Leu Pro Ser Ala Val Val Ala	
-21 -20 -15 -10	
GCC CTG CCG GTG TTG GCC CTT GCC GCT GAT GGC AGG TCC ACC CGC TAC	96
Ala Leu Pro Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr	
-5 1 5	
TGG GAC TGC TGC AAG CCT TCG TGC GGC TGG GCC AAG AAG GCT CCC GTG	144
Trp Asp Cys Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val	
10 15 20	
AAC CAG CCT GTC TTT TCC TGC AAC GCC AAC TTC CAG CGT ATC ACG GAC	192
Asn Gln Pro Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp	
25 30 35 40	
TTC GAC GCC AAG TCC GGC TGC GAG CCG GGC GGT GTC GCC TAC TCG TGC	240
Phe Asp Ala Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys	
45 50 55	
GCC GAC CAG ACC CCA TGG GCT GTG AAC GAC GAC TTC GCG CTC GGT TTT	288
Ala Asp Gln Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe	
60 65 70	
GCT GCC ACC TCT ATT GCC GGC AGC AAT GAG GCG GGC TGG TGC TGC GCC	336
Ala Ala Thr Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala	
75 80 85	
TGC TAC GAG CTC ACC TTC ACA TCC GGT CCT GTT GCT GGC AAG AAG ATG	384
Cys Tyr Glu Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met	
90 95 100	
GTC GTC CAG TCC ACC AGC ACT GGC GGT GAT CTT GGC AGC AAC CAC TTC	432
Val Val Gln Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe	
105 110 115 120	
GAT CTC AAC ATC CCC GGC GGC GGC GTC GGC ATC TTC GAC GGA TGC ACT	480

Asp	Leu	Asn	Ile	Pro	Gly	Gly	Gly	Val	Gly	Ile	Phe	Asp	Gly	Cys	Thr		
				125					130					135			
CCC	CAG	TTC	GGC	GGT	CTG	CCC	GGC	CAG	CGC	TAC	GGC	GGC	ATC	TCG	TCC	528	
Pro	Gln	Phe	Gly	Gly	Leu	Pro	Gly	Gln	Arg	Tyr	Gly	Gly	Ile	Ser	Ser		
			140					145					150				
CGC	AAC	GAG	TGC	GAT	CGG	TTC	CCC	GAC	GCC	CTC	AAG	CCC	GGC	TGC	TAC	576	
Arg	Asn	Glu	Cys	Asp	Arg	Phe	Pro	Asp	Ala	Leu	Lys	Pro	Gly	Cys	Tyr		
		155					160					165					
TGG	CGC	TTC	GAC	TGG	TTC	AAG	AAC	GCC	GAC	AAT	CCG	AGC	TTC	AGC	TTC	624	
Trp	Arg	Phe	Asp	Trp	Phe	Lys	Asn	Ala	Asp	Asn	Pro	Ser	Phe	Ser	Phe		
	170					175				180							
CGT	CAG	GTC	CAG	TGC	CCA	GCC	GAG	CTC	GTC	GCT	CGC	ACC	GGA	TGC	CGC	672	
Arg	Gln	Val	Gln	Cys	Pro	Ala	Glu	Leu	Val	Ala	Arg	Thr	Gly	Cys	Arg		
185					190				195				200				
CGC	AAC	GAC	GAC	GGC	AAC	TTC	CCT	GCC	GTC	CAG	ATC	CCC	TCC	AGC	AGC	720	
Arg	Asn	Asp	Asp	Gly	Asn	Phe	Pro	Ala	Val	Gln	Ile	Pro	Ser	Ser	Ser		
				205				210					215				
ACC	AGC	TCT	CCG	GTC	AAC	CAG	CCT	ACC	AGC	ACC	AGC	ACC	ACG	TCC	ACC	768	
Thr	Ser	Ser	Pro	Val	Asn	Gln	Pro	Thr	Ser	Thr	Ser	Thr	Thr	Ser	Thr		
			220				225					230					
TCC	ACC	ACC	TCG	AGC	CCG	CCA	GTC	CAG	CCT	ACG	ACT	CCC	AGC	GGC	TGC	816	
Ser	Thr	Thr	Ser	Ser	Pro	Pro	Val	Gln	Pro	Thr	Thr	Pro	Ser	Gly	Cys		
			235				240					245					
ACT	GCT	GAG	AGG	TGG	GCT	CAG	TGC	GGC	GGC	AAT	GGC	TGG	AGC	GGC	TGC	864	
Thr	Ala	Glu	Arg	Trp	Ala	Gln	Cys	Gly	Gly	Asn	Gly	Trp	Ser	Gly	Cys		
	250					255				260							
ACC	ACC	TGC	GTC	GCT	GGC	AGC	ACT	TGC	ACG	AAG	ATT	AAT	GAC	TGG	TAC	912	
Thr	Thr	Cys	Val	Ala	Gly	Ser	Thr	Cys	Thr	Lys	Ile	Asn	Asp	Trp	Tyr		
265					270				275					280			
CAT	CAG	TGC	CTG	TAGACGCAGG	GCAGCTTGAG	GGCCTTACTG	GTGGCCGCAA									964	
His	Gln	Cys	Leu														
CGAAATGACA	CTCCCAATCA	CTGTATTAGT	TCTTGACAT	AATTCGTCA	TCCCTCCAGG											1024	
GATTGTCACA	TAAATGCAAT	GAGGAACAAT	GAGTAC													1060	

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 305 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Arg Ser Ser Pro Leu Leu Pro Ser Ala Val Val Ala Ala Leu Pro
 -21 -20 -15 -10
 Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr Trp Asp Cys
 -5 1 5 10
 Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val Asn Gln Pro
 15 20 25
 Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp Phe Asp Ala
 30 35 40
 Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys Ala Asp Gln
 45 50 55
 Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe Ala Ala Thr
 60 65 70 75
 Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala Cys Tyr Glu
 80 85 90
 Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met Val Val Gln
 95 100 105
 Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe Asp Leu Asn
 110 115 120
 Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Pro Gln Phe
 125 130 135
 Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser Arg Asn Glu
 140 145 150 155
 Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr Trp Arg Phe
 160 165 170
 Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe Arg Gln Val
 175 180 185
 Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg Arg Asn Asp
 190 195 200
 Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Ser Ser Ser Thr Ser Ser
 205 210 215
 Pro Val Asn Gln Pro Thr Ser Thr Ser Thr Ser Thr Ser Thr Thr
 220 225 230 235
 Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys Thr Ala Glu
 240 245 250
 Arg Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys Thr Thr Cys
 255 260 265

Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr His Gln Cys
 270 275 280

Leu

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1473 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Fusarium oxysporum*
- (B) STRAIN: DSM 2672

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 97..1224

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GAATTCGCGG CCGCTCATTC ACTTCATTCA TTCTTTAGAA TTACATACAC TCTCTTTCAA	60
AACAGTCACT CTTTAAACAA AACAACTTTT GCAACA ATG CGA TCT TAC ACT CTT	114
Met Arg Ser Tyr Thr Leu	
1 5	
CTC GCC CTG GCC GGC CCT CTC GCC GTG AGT GCT GCT TCT GGA AGC GGT	162
Leu Ala Leu Ala Gly Pro Leu Ala Val Ser Ala Ala Ser Gly Ser Gly	
10 15 20	
CAC TCT ACT CGA TAC TGG GAT TGC TGC AAG CCT TCT TGC TCT TGG AGC	210
His Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser Cys Ser Trp Ser	
25 30 35	
GGA AAG GCT GCT GTC AAC GCC CCT GCT TTA ACT TGT GAT AAG AAC GAC	258
Gly Lys Ala Ala Val Asn Ala Pro Ala Leu Thr Cys Asp Lys Asn Asp	
40 45 50	
AAC CCC ATT TCC AAC ACC AAT GCT GTC AAC GGT TGT GAG GGT GGT GGT	306
Asn Pro Ile Ser Asn Thr Asn Ala Val Asn Gly Cys Glu Gly Gly Gly	
55 60 65 70	
TCT GCT TAT GCT TGC ACC AAC TAC TCT CCC TGG GCT GTC AAC GAT GAG	354
Ser Ala Tyr Ala Cys Thr Asn Tyr Ser Pro Trp Ala Val Asn Asp Glu	
75 80 85	

CTT GCC TAC GGT TTC GCT GCT ACC AAG ATC TCC GGT GGC TCC GAG GCC	402
Leu Ala Tyr Gly Phe Ala Ala Thr Lys Ile Ser Gly Gly Ser Glu Ala	
90 95 100	
AGC TGG TGC TGT GCT TGC TAT GCT TTG ACC TTC ACC ACT GGC CCC GTC	450
Ser Trp Cys Cys Ala Cys Tyr Ala Leu Thr Phe Thr Thr Gly Pro Val	
105 110 115	
AAG GGC AAG AAG ATG ATC GTC CAG TCC ACC AAC ACT GGA GGT GAT CTC	498
Lys Gly Lys Lys Met Ile Val Gln Ser Thr Asn Thr Gly Gly Asp Leu	
120 125 130	
GGC GAC AAC CAC TTC GAT CTC ATG ATG CCC GGC GGT GGT GTC GGT ATC	546
Gly Asp Asn His Phe Asp Leu Met Met Pro Gly Gly Gly Val Gly Ile	
135 140 145 150	
TTC GAC GGC TGC ACC TCT GAG TTC GGC AAG GCT CTC GGC GGT GCC CAG	594
Phe Asp Gly Cys Thr Ser Glu Phe Gly Lys Ala Leu Gly Gly Ala Gln	
155 160 165	
TAC GGC GGT ATC TCC TCC CGA AGC GAA TGT GAT AGC TAC CCC GAG CTT	642
Tyr Gly Gly Ile Ser Ser Arg Ser Glu Cys Asp Ser Tyr Pro Glu Leu	
170 175 180	
CTC AAG GAC GGT TGC CAC TGG CGA TTC GAC TGG TTC GAG AAC GCC GAC	690
Leu Lys Asp Gly Cys His Trp Arg Phe Asp Trp Phe Glu Asn Ala Asp	
185 190 195	
AAC CCT GAC TTC ACC TTT GAG CAG GTT CAG TGC CCC AAG GCT CTC CTC	738
Asn Pro Asp Phe Thr Phe Glu Gln Val Gln Cys Pro Lys Ala Leu Leu	
200 205 210	
GAC ATC AGT GGA TGC AAG CGT GAT GAC GAC TCC AGC TTC CCT GCC TTC	786
Asp Ile Ser Gly Cys Lys Arg Asp Asp Asp Ser Ser Phe Pro Ala Phe	
215 220 225 230	
AAG GTT GAT ACC TCG GCC AGC AAG CCC CAG CCC TCC AGC TCC GCT AAG	834
Lys Val Asp Thr Ser Ala Ser Lys Pro Gln Pro Ser Ser Ser Ala Lys	
235 240 245	
AAG ACC ACC TCC GCT GCT GCT GCC GCT CAG CCC CAG AAG ACC AAG GAT	882
Lys Thr Thr Ser Ala Ala Ala Ala Gln Pro Gln Lys Thr Lys Asp	
250 255 260	
TCC GCT CCT GTT GTC CAG AAG TCC TCC ACC AAG CCT GCC GCT CAG CCC	930
Ser Ala Pro Val Val Gln Lys Ser Ser Thr Lys Pro Ala Ala Gln Pro	
265 270 275	
GAG CCT ACT AAG CCC GCC GAC AAG CCC CAG ACC GAC AAG CCT GTC GCC	978
Glu Pro Thr Lys Pro Ala Asp Lys Pro Gln Thr Asp Lys Pro Val Ala	
280 285 290	
ACC AAG CCT GCT GCT ACC AAG CCC GTC CAA CCT GTC AAC AAG CCC AAG	1026
Thr Lys Pro Ala Ala Thr Lys Pro Val Gln Pro Val Asn Lys Pro Lys	
295 300 305 310	

ACA ACC CAG AAG GTC CGT GGA ACC AAA ACC CGA GGA AGC TGC CCG GCC	1074
Thr Thr Gln Lys Val Arg Gly Thr Lys Thr Arg Gly Ser Cys Pro Ala	
315 320 325	
AAG ACT GAC GCT ACC GCC AAG GCC TCC GTT GTC CCT GCT TAT TAC CAG	1122
Lys Thr Asp Ala Thr Ala Lys Ala Ser Val Val Pro Ala Tyr Tyr Gln	
330 335 340	
TGT GGT GGT TCC AAG TCC GCT TAT CCC AAC GGC AAC CTC GCT TGC GCT	1170
Cys Gly Gly Ser Lys Ser Ala Tyr Pro Asn Gly Asn Leu Ala Cys Ala	
345 350 355	
ACT GGA AGC AAG TGT GTC AAG CAG AAC GAG TAC TAC TCC CAG TGT GTC	1218
Thr Gly Ser Lys Cys Val Lys Gln Asn Glu Tyr Tyr Ser Gln Cys Val	
360 365 370	
CCC AAC TAAATGGTAG ATCCATCGGT TGTGGAAGAG ACTATGCGTC TCAGAAGGGA	1274
Pro Asn	
375	
TCCTCTCATG AGCAGGCTTG TCATTGTATA GCATGGCATC CTGGACCAAG TGTTCGACCC	1334
TTGTTGTACA TAGTATATCT TCATTGTATA TATTTAGACA CATAGATAGC CTCTTGTCAG	1394
CGACAACTGG CTACAAAAGA CTTGGCAGGC TTGTTCAATA TTGACACAGT TTCCTCCATA	1454
AAAAAAAAA AAAAAAAAAA	1473

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 376 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Ser Tyr Thr Leu Leu Ala Leu Ala Gly Pro Leu Ala Val Ser	
1 5 10 15	
Ala Ala Ser Gly Ser Gly His Ser Thr Arg Tyr Trp Asp Cys Cys Lys	
20 25 30	
Pro Ser Cys Ser Trp Ser Gly Lys Ala Ala Val Asn Ala Pro Ala Leu	
35 40 45	
Thr Cys Asp Lys Asn Asp Asn Pro Ile Ser Asn Thr Asn Ala Val Asn	
50 55 60	
Gly Cys Glu Gly Gly Gly Ser Ala Tyr Ala Cys Thr Asn Tyr Ser Pro	
65 70 75 80	
Trp Ala Val Asn Asp Glu Leu Ala Tyr Gly Phe Ala Ala Thr Lys Ile	

54

85

90

95

Ser Gly Gly Ser Glu Ala Ser Trp Cys Cys Ala Cys Tyr Ala Leu Thr
 100 105 110

Phe Thr Thr Gly Pro Val Lys Gly Lys Lys Met Ile Val Gln Ser Thr
 115 120 125

Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Met Met Pro
 130 135 140

Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Ser Glu Phe Gly Lys
 145 150 155 160

Ala Leu Gly Gly Ala Gln Tyr Gly Gly Ile Ser Ser Arg Ser Glu Cys
 165 170 175

Asp Ser Tyr Pro Glu Leu Leu Lys Asp Gly Cys His Trp Arg Phe Asp
 180 185 190

Trp Phe Glu Asn Ala Asp Asn Pro Asp Phe Thr Phe Glu Gln Val Gln
 195 200 205

Cys Pro Lys Ala Leu Leu Asp Ile Ser Gly Cys Lys Arg Asp Asp Asp
 210 215 220

Ser Ser Phe Pro Ala Phe Lys Val Asp Thr Ser Ala Ser Lys Pro Gln
 225 230 235 240

Pro Ser Ser Ser Ala Lys Lys Thr Thr Ser Ala Ala Ala Ala Ala Gln
 245 250 255

Pro Gln Lys Thr Lys Asp Ser Ala Pro Val Val Gln Lys Ser Ser Thr
 260 265 270

Lys Pro Ala Ala Gln Pro Glu Pro Thr Lys Pro Ala Asp Lys Pro Gln
 275 280 285

Thr Asp Lys Pro Val Ala Thr Lys Pro Ala Ala Thr Lys Pro Val Gln
 290 295 300

Pro Val Asn Lys Pro Lys Thr Thr Gln Lys Val Arg Gly Thr Lys Thr
 305 310 315 320

Arg Gly Ser Cys Pro Ala Lys Thr Asp Ala Thr Ala Lys Ala Ser Val
 325 330 335

Val Pro Ala Tyr Tyr Gln Cys Gly Gly Ser Lys Ser Ala Tyr Pro Asn
 340 345 350

Gly Asn Leu Ala Cys Ala Thr Gly Ser Lys Cys Val Lys Gln Asn Glu
 355 360 365

Tyr Tyr Ser Gln Cys Val Pro Asn
 370 375

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGCTGCGGCC GCAGGCCGCG GAGGCCA

27

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGCTTGGCCT CCGCGGCCTG CGGCCGC

27

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AATTCGCGGC CGCGCCATG GAGGCC

26

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AATTGGCCTC CATGGCCGCG GCCGCG

26

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 15 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AAYGCGACA AAYCC

15

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AACGAYGAYG GNAAYTTCCC

20

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

AAYGAYTGGT ACCAYCARTG

20

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GCGCCAGTAG CAGCCGGGCT TGAGGG

26

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

ACGTCTCAAC TCGGATCCAA GATGCGTT

28

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 25 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CTCAACTCTG ATCAAGATGC GTTCC

25

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 26 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TGTCGACCAG TAAGGCCCTC AAGCTG

26

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GACAGAGCAC AGAATTCAC AGTGAGCTCT

30

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TGGGAYTGYT GYAARCC

17

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

AGGGAGACCG GAATTCTGGG AYTGYTGYAA RCC

33

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CCNGGNGGNG GNGTNNG

17

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGGGAGACCG GAATCCCNG GNGGNGNGT NGG

33

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACNAYCATNK TYTTNCC

17

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 34 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GACAGAGCAC AGAATTCACN AYCATTNKTYT TNCC

34

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

NGGRTRTCN GCNKYYTYRA ACCA

24

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 41 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GACAGAGCAC AGAATTCNGG RTTRTCNGCN KYYTYRAACC A

41

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGGGTAGCTA TCACATTCGC TTCGGGAGGA GATACCGCCG TA

42

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CTTCTTGCTC TTGGAGCGGA AAGGCTGCTG TCAACGCCCC TG

42

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TGTACGCATG TAACATTA

18

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTGCACAATA TTTCAAGC

18

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGGGTAGCTA TCACATTCGC TTCGGGAGGA GATACCGCCG TA

42

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 42 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTTCTTGCTC TTGGAGCGGA AAGGCTGCTG TCAACGCCCC TG

42

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AGCTTCTCAA GGACGGTT

18

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

AACAAGGGTC GAACACTT

18

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

CCAGAAGACC AAGGATT

17

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- 5 (i) APPLICANT: NOVO NORDISK A/S, N N
- (ii) TITLE OF INVENTION: A Cellulase Preparation
- 10 (iii) NUMBER OF SEQUENCES: 4
- (iv) CORRESPONDENCE ADDRESS:
- 15 (A) ADDRESSEE: NOVO NORDISK A/S, Patent Department
- (B) STREET: Novo Alle
- (C) CITY: Bagsvaerd
- (E) COUNTRY: DENMARK
- (F) ZIP: DK-2880
- (v) COMPUTER READABLE FORM:
- 20 (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- 25 (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:
- 30 (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Thalsoe-Madsen, Birgit
- (ix) TELECOMMUNICATION INFORMATION:
- 35 (A) TELEPHONE: +45 4444 8888
- (B) TELEFAX: +45 4449 3256
- (C) TELEX: 37304

(2) INFORMATION FOR SEQ ID NO:1:

- 40 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1060 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- 45 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA
- (iii) HYPOTHETICAL: NO
- 50 (vi) ORIGINAL SOURCE:
- (A) ORGANISM: Humicola insolens
- (B) STRAIN: DSM 1800

(ix) FEATURE:

(A) NAME/KEY: mat_peptide
(B) LOCATION: 73..927

5

(ix) FEATURE:

(A) NAME/KEY: sig_peptide
(B) LOCATION: 10..72

10

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 10..927

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

15

GGATCCAAG ATG OGT TCC TCC CCC CTC CTC CCG TCC GCC GGT GIG GCC 48
Met Arg Ser Ser Pro Leu Leu Pro Ser Ala Val Val Ala
-21 -20 -15 -10

20

GCC CTG CCG GTG TTG GCC CTT GCC GCT GAT GGC AGG TCC ACC CGC TAC 96
Ala Leu Pro Val Leu Ala Leu Ala Ala Asp Gly Arg Ser Thr Arg Tyr
-5 1 5

25

TGG GAC TGC TGC AAG CCT TOG TGC GGC TGG GCC AAG AAG GCT CCC GTG 144
Trp Asp Cys Cys Lys Pro Ser Cys Gly Trp Ala Lys Lys Ala Pro Val
10 15 20

30

AAC CAG CCT GTC TTT TCC TGC AAC GGC AAC TTC CAG CGT ATC ACG GAC 192
Asn Gln Pro Val Phe Ser Cys Asn Ala Asn Phe Gln Arg Ile Thr Asp
25 30 35 40

35

TTC GAC GCC AAG TCC GGC TGC GAG CCG GGC GGT GTC GCC TAC TCG TGC 240
Phe Asp Ala Lys Ser Gly Cys Glu Pro Gly Gly Val Ala Tyr Ser Cys
45 50 55

40

GCC GAC CAG ACC CCA TGG GCT GTG AAC GAC GAC TTC GCG CTC GGT TTT 288
Ala Asp Gln Thr Pro Trp Ala Val Asn Asp Asp Phe Ala Leu Gly Phe
60 65 70

45

GCT GCC ACC TCT ATT GCC GGC AGC AAT GAG GCG GGC TGG TGC TGC GCC 336
Ala Ala Thr Ser Ile Ala Gly Ser Asn Glu Ala Gly Trp Cys Cys Ala
75 80 85

50

TGC TAC GAG CTC ACC TTC ACA TCC GGT CCT GGT GCT GGC AAG AAG ATG 384
Cys Tyr Glu Leu Thr Phe Thr Ser Gly Pro Val Ala Gly Lys Lys Met
90 95 100

55

GTC GTC CAG TCC ACC AGC ACT GGC GGT GAT CTT GGC AGC AAC CAC TTC 432
Val Val Gln Ser Thr Ser Thr Gly Gly Asp Leu Gly Ser Asn His Phe
105 110 115 120

60

GAT CTC AAC ATC CCC GGC GGC GGC GTC GGC ATC TTC GAC GGA TGC ACT 480
Asp Leu Asn Ile Pro Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr
125 130 135

55

	CCC CAG TTC GGC GGT CTG CCC GGC CAG GGC TAC GGC GGC ATC TCG TCC	528
	Pro Gln Phe Gly Gly Leu Pro Gly Gln Arg Tyr Gly Gly Ile Ser Ser	
	140 145 150	
5	CGC AAC GAG TGC GAT GGG TTC CCC GAC GGC CTC AAG CCC GGC TGC TAC	576
	Arg Asn Glu Cys Asp Arg Phe Pro Asp Ala Leu Lys Pro Gly Cys Tyr	
	155 160 165	
10	TGG CGC TTC GAC TGG TTC AAG AAC GGC GAC AAT CCG AGC TTC AGC TTC	624
	Trp Arg Phe Asp Trp Phe Lys Asn Ala Asp Asn Pro Ser Phe Ser Phe	
	170 175 180	
15	CGT CAG GTC CAG TGC CCA GGC GAG CTC GTC GCT CGC ACC GGA TGC CGC	672
	Arg Gln Val Gln Cys Pro Ala Glu Leu Val Ala Arg Thr Gly Cys Arg	
	185 190 195 200	
20	CGC AAC GAC GAC GGC AAC TTC CCT GCC GTC CAG ATC CCC TCC AGC AGC	720
	Arg Asn Asp Asp Gly Asn Phe Pro Ala Val Gln Ile Pro Ser Ser Ser	
	205 210 215	
25	ACC AGC TCT CCG GTC AAC CAG CCT ACC AGC ACC AGC ACC ACG TCC ACC	768
	Thr Ser Ser Pro Val Asn Gln Pro Thr Ser Thr Ser Thr Thr Ser Thr	
	220 225 230	
30	TCC ACC ACC TCG AGC CCG CCA GTC CAG CCT ACG ACT CCC AGC GGC TGC	816
	Ser Thr Thr Ser Ser Pro Pro Val Gln Pro Thr Thr Pro Ser Gly Cys	
	235 240 245	
35	ACT GCT GAG AGG TGG GCT CAG TGC GGC GGC AAT GGC TGG AGC GGC TGC	864
	Thr Ala Glu Arg Trp Ala Gln Cys Gly Gly Asn Gly Trp Ser Gly Cys	
	250 255 260	
40	ACC ACC TGC GTC GCT GGC AGC ACT TGC ACG AAG ATT AAT GAC TGG TAC	912
	Thr Thr Cys Val Ala Gly Ser Thr Cys Thr Lys Ile Asn Asp Trp Tyr	
	265 270 275 280	
45	CAT CAG TGC CTG TAGACGCAGG GCAGCTTGAG GGCCCTTACAG GTGGCCGCAA	964
	His Gln Cys Leu	
	285	
50	CGAAATGACA CTCCTAATCA CIGIATTAGT TCTTGTTACAT AATTTCGICA TCCCTCCAGG	1024
	GATTGTGACA TAAATGCAAT GAGGAACAAT GAGTAC	1060

[illegible]

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 1473 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

10

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

15

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: *Fusarium oxysporum*
 (B) STRAIN: DSM 2672

(ix) FEATURE:

20

- (A) NAME/KEY: CDS
 (B) LOCATION: 97..1224

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

25

GAATTCGGGG COGCTCATTC ACTTCATICA TTCTTTAGAA TTACATACAC TCCTTTTCAA 60

AACAGTCACT CTTTAAACAA AACAACTTTT GCAACA ATG CGA TCT TAC ACT CTT 114

Met Arg Ser Tyr Thr Leu
 1 5

30

CTC GCC CTG GCC GGC CCT CTC GCC GIG AGT GCT GCT TCT GGA AGC GGT 162

Leu Ala Leu Ala Gly Pro Leu Ala Val Ser Ala Ala Ser Gly Ser Gly
 10 15 20

35

CAC TCT ACT CGA TAC TGG GAT TGC TGC AAG CCT TCT TGC TCT TGG AGC 210

His Ser Thr Arg Tyr Trp Asp Cys Cys Lys Pro Ser Cys Ser Trp Ser
 25 30 35

40

GGA AAG GCT GCT GTC AAC GCC CCT GCT TTA ACT TGT GAT AAG AAC GAC 258

Gly Lys Ala Ala Val Asn Ala Pro Ala Leu Thr Cys Asp Lys Asn Asp
 40 45 50

45

AAC CCC ATT TCC AAC ACC AAT GCT GTC AAC GGT TGT GAG GGT GGT GGT 306

Asn Pro Ile Ser Asn Thr Asn Ala Val Asn Gly Cys Glu Gly Gly Gly
 55 60 65 70

50

TCT GCT TAT GCT TGC ACC AAC TAC TCT CCC TGG GCT GTC AAC GAT GAG 354

Ser Ala Tyr Ala Cys Thr Asn Tyr Ser Pro Trp Ala Val Asn Asp Glu
 75 80 85

55

CTT GCC TAC GGT TTC GCT GCT ACC AAG ATC TCC GGT GGC TCC GAG GCC 402

Leu Ala Tyr Gly Phe Ala Ala Thr Lys Ile Ser Gly Gly Ser Glu Ala
 90 95 100

	AGC TGG TGC TGT GCT TGC TAT GCT TTG ACC TTC ACC ACT GGC CCC GTC Ser Trp Cys Cys Ala Cys Tyr Ala Leu Thr Phe Thr Thr Gly Pro Val 105 110 115	450
5	AAG GGC AAG AAG ATG ATC GTC CAG TCC ACC AAC ACT GGA GGT GAT CTC Lys Gly Lys Lys Met Ile Val Gln Ser Thr Asn Thr Gly Gly Asp Leu 120 125 130	498
10	GGC GAC AAC CAC TTC GAT CTC ATG ATG CCC GGC GGT GGT GTC GGT ATC Gly Asp Asn His Phe Asp Leu Met Met Pro Gly Gly Gly Val Gly Ile 135 140 145 150	546
15	TTC GAC GGC TGC ACC TCT GAG TTC GGC AAG GCT CTC GGC GGT GCC CAG Phe Asp Gly Cys Thr Ser Glu Phe Gly Lys Ala Leu Gly Gly Ala Gln 155 160 165	594
20	TAC GGC GGT ATC TCC TCC CGA AGC GAA TGT GAT AGC TAC CCC GAG CTT Tyr Gly Gly Ile Ser Ser Arg Ser Glu Cys Asp Ser Tyr Pro Glu Leu 170 175 180	642
	CTC AAG GAC GGT TGC CAC TGG CGA TTC GAC TGG TTC GAG AAC GCC GAC Leu Lys Asp Gly Cys His Trp Arg Phe Asp Trp Phe Glu Asn Ala Asp 185 190 195	690
25	AAC CCT GAC TTC ACC TTT GAG CAG GTT CAG TGC CCC AAG GCT CTC CTC Asn Pro Asp Phe Thr Phe Glu Gln Val Gln Cys Pro Lys Ala Leu Leu 200 205 210	738
30	GAC ATC AGT GGA TGC AAG CGT GAT GAC GAC TCC AGC TTC CCT GCC TTC Asp Ile Ser Gly Cys Lys Arg Asp Asp Asp Ser Ser Phe Pro Ala Phe 215 220 225 230	786
35	AAG GTT GAT ACC TCG GCC AGC AAG CCC CAG CCC TCC AGC TCC GCT AAG Lys Val Asp Thr Ser Ala Ser Lys Pro Gln Pro Ser Ser Ser Ala Lys 235 240 245	834
40	AAG ACC ACC TCC GCT GCT GCT GCC GCT CAG CCC CAG AAG ACC AAG GAT Lys Thr Thr Ser Ala Ala Ala Ala Ala Gln Pro Gln Lys Thr Lys Asp 250 255 260	882
	TCC GCT CCT GTT GTC CAG AAG TCC TCC ACC AAG CCT GCC GCT CAG CCC Ser Ala Pro Val Val Gln Lys Ser Ser Thr Lys Pro Ala Ala Gln Pro 265 270 275	930
45	GAG CCT ACT AAG CCC GCC GAC AAG CCC CAG ACC GAC AAG CCT GTC GCC Glu Pro Thr Lys Pro Ala Asp Lys Pro Gln Thr Asp Lys Pro Val Ala 280 285 290	978
50	ACC AAG CCT GCT GCT ACC AAG CCC GTC CAA CCT GTC AAC AAG CCC AAG Thr Lys Pro Ala Ala Thr Lys Pro Val Gln Pro Val Asn Lys Pro Lys 295 300 305 310	1026
55	ACA ACC CAG AAG GTC CGT GGA ACC AAA ACC CGA GGA ACC TGC CCG GCC Thr Thr Gln Lys Val Arg Gly Thr Lys Thr Arg Gly Ser Cys Pro Ala 315 320 325	1074

	AAG ACT GAC GCT ACC GGC AAG GGC TCC GGT GTC OCT GCT TAT TAC CAG	1122
	Lys Thr Asp Ala Thr Ala Lys Ala Ser Val Val Pro Ala Tyr Tyr Gln	
	330 335 340	
5	TGT GGT GGT TCC AAG TCC GCT TAT CCC AAC GGC AAC CTC GCT TGC GCT	1170
	Cys Gly Gly Ser Lys Ser Ala Tyr Pro Asn Gly Asn Leu Ala Cys Ala	
	345 350 355	
10	ACT GGA AGC AAG TGT GTC AAG CAG AAC GAG TAC TAC TCC CAG TGT GTC	1218
	Thr Gly Ser Lys Cys Val Lys Gln Asn Glu Tyr Tyr Ser Gln Cys Val	
	360 365 370	
	CCC AAC TAAATGGTAG ATCCATCGGT TGTGGAAGAG ACTATGGGTC TCAGAAGGCA	1274
	Pro Asn	
15	375	
	TCCTCTCATG AGCAGGCTTG TCATTGTATA GCATGGCATC CIGGACCAAG TGTTCGACCC	1334
	TGTGTGTACA TAGTATATCT TCATTGTATA TATTTAGACA CATAGATAGC CTCGTGTGAG	1394
20	CGACAACCTGG CTACAAAAGA CTGGGCAGGC TTGTTCATAA TTGACACAGT TTCTTCATA	1454
	AAAAAAAAA AAAAAAAAAA	1473

[illegible]

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- 5 (A) LENGTH: 376 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Arg Ser Tyr Thr Leu Leu Ala Leu Ala Gly Pro Leu Ala Val Ser
1 5 10 15
Ala Ala Ser Gly Ser Gly His Ser Thr Arg Tyr Trp Asp Cys Cys Lys
20 25 30
Pro Ser Cys Ser Trp Ser Gly Lys Ala Ala Val Asn Ala Pro Ala Leu
20 35 40 45
Thr Cys Asp Lys Asn Asp Asn Pro Ile Ser Asn Thr Asn Ala Val Asn
50 55 60
25 Gly Cys Glu Gly Gly Gly Ser Ala Tyr Ala Cys Thr Asn Tyr Ser Pro
65 70 75 80
Trp Ala Val Asn Asp Glu Leu Ala Tyr Gly Phe Ala Ala Thr Lys Ile
85 90 95
30 Ser Gly Gly Ser Glu Ala Ser Trp Cys Cys Ala Cys Tyr Ala Leu Thr
100 105 110
Phe Thr Thr Gly Pro Val Lys Gly Lys Lys Met Ile Val Gln Ser Thr
35 115 120 125
Asn Thr Gly Gly Asp Leu Gly Asp Asn His Phe Asp Leu Met Met Pro
130 135 140
40 Gly Gly Gly Val Gly Ile Phe Asp Gly Cys Thr Ser Glu Phe Gly Lys
145 150 155 160
Ala Leu Gly Gly Ala Gln Tyr Gly Gly Ile Ser Ser Arg Ser Glu Cys
165 170 175
45 Asp Ser Tyr Pro Glu Leu Leu Lys Asp Gly Cys His Trp Arg Phe Asp
180 185 190
Trp Phe Glu Asn Ala Asp Asn Pro Asp Phe Thr Phe Glu Gln Val Gln
50 195 200 205
Cys Pro Lys Ala Leu Leu Asp Ile Ser Gly Cys Lys Arg Asp Asp Asp
210 215 220
55 Ser Ser Phe Pro Ala Phe Lys Val Asp Thr Ser Ala Ser Lys Pro Gln
225 230 235 240

00735787-14300

